

PATENT APPLICATION

Logical Unit Security for Clustered Storage Area Networks

Inventor: **Yuichi Taguchi**
Residence: San Jose, California
Citizenship: Japan

Assignee: **Hitachi, Ltd.**
6, Kanda Surugadai 4-chome
Chiyoda-ku, Tokyo, Japan
Incorporation: Japan

Entity: Large

Logical Unit Security for Clustered Storage Area Networks

BACKGROUND OF THE INVENTION

[0001] This invention relates to storage area networks, and in particular, to security of data in such storage area networks.

[0002] Storage area networks (SAN) are now widely deployed around the world to provide storage facilities for computing environments. The consolidation of storage into large facilities allows for more efficient administration and control, and enables the provision of highly reliable backup and redundancy systems to dramatically limit data loss from system failures or natural catastrophes. The benefits of such storage have caused a dramatic increase in its use throughout the world.

[0003] Storage area networks are designed to allow access from many different host systems so the data may be reliably retrieved and stored from different locations under the control of different processors. Such storage and retrieval is often carried out over public networks, such as the internet.

[0004] To more effectively enable access to the storage system, it is usually configured into smaller portions which can be allocated to different processors or other resources requiring the storage. For example, conventional storage systems include a large number of hard disk drives, with each hard disk drive itself including many gigabytes of storage. One way to divide the storage resource into smaller portions is to create Logical Units that are assigned a unique Logical Unit Number. Each LU itself consists of a numeric address, thereby permitting the large storage system to appear as many smaller storage units to the host computers accessing them, enabling more efficient operation.

[0005] The LUs are typically assigned to hosts so that a particular LU may be accessed only by designated hosts which have "permission" to access that portion of the storage system. This provides enhanced security as the software controlling access to the storage system will only permit access by certain previously defined hosts. While it is somewhat arbitrary how many hosts are permitted to access a given LU, conventionally only one host usually has access rights to a given LU at a given time. In this manner the data on the LU is protected against access by hosts or servers other than those previously designated, thereby enhancing the security of the data.

[0006] The SAN itself usually consists of one or more disk array devices, for example configured under a selected RAID protocol, with multiple host devices connected by fiber channel switch network devices or other well known means to the storage area network. Inside the architecture, host computers run cluster management software to negotiate with each other to determine which hosts "own" which portions of the storage or LUs. A commonly known "failover" function enables the clustered architecture of the SAN to be highly available. In a failover situation, a failure occurs, but is made relatively transparent to the user of the system by transferring operations that were running on one node to another node or nodes within that cluster of nodes.

[0007] The hosts and SAN are configured in a clustered environment in which more than one host has access to a given SAN. The host is typically a computer, for example an application server or a database server. The traditional clustering software used to provide this configuration has a limitation in that a portion of the storage, for example, an LUN, must be configured to allow I/O access from all host nodes in the cluster. For example, if host node A is running an application X, and they are in the same cluster, it is desirable that if host A fails, the task of application X be taken over by another host B. When this happens the LU security for applications must be set to allow I/O access from both hosts A and B. If such multiple access is routinely provided by the set-up program at system initialization, however, it can be a cause of data corruption resulting from the wrong access of the data.

[0008] This invention provides an improved method and system for security in such an environment by enabling dynamic changes in the LU security to allow a different host to access a particular portion of the storage after a failure, when that host could not have accessed that portion of the storage before the failure.

BRIEF SUMMARY OF THE INVENTION

[0009] The system of this invention is particularly applicable to host nodes which are configured in a cluster system. During the initialization of such a system, the host computers negotiate to gain ownership to access the data storage resources which are divided on the basis of LUs (or volumes or other units). Each LU is then configured to permit and reject I/O accesses by hosts, typically based on IDs assigned to the hosts, such as WWN, port ID. In the primary security configuration, one LU may allow I/O access only from one host.

[0010] According to this invention, if a problem occurs in the primary host group (assigned to the particular LU), another host group takes over execution of that process to continue system activity. At the time this occurs, the cluster management software running on the host detects the change in ownership and notifies the storage device that ownership for that LU has been changed. The LU security function in the storage device then dynamically changes its security configuration to allow I/O accesses from the new host. As a result, this invention provides improved security. Secondary hosts within the cluster that run the application processes are thereby not permitted to access the LU until they have obtained that right, typically as a result of a failure in the primary host. While operating in this manner, data corruption at the LU level is greatly reduced, and access of that data by unauthorized applications is generally prevented.

[0011] In one embodiment in a system having a first host computer, a second host computer, and a storage, and in which access to at least a portion of the storage is controlled to permit access to the storage by the first host, and prevent access to the storage by the second host, a method of changing the authorization for access to the storage includes informing the security system of a negotiation between the hosts and restricting access to the portions to the host that acquired such access in the negotiation.

BRIEF DESCRIPTION OF THE DRAWINGS

- [0012] Figure 1 is a block diagram illustrating an in-band system implementation;
- [0013] Figure 2 is a block diagram illustrating an out-of-band system implementation;
- [0014] Figure 3 illustrates a sample data structure for storage configuration information;
- [0015] Figure 4 illustrates a sample data structure for security configuration information;
- [0016] Figure 5 is a flowchart illustrating one embodiment of the method of this invention;
- [0017] Figure 6 is a diagram illustrating an ownership change message; and
- [0018] Figure 7 is a diagram illustrating an access control status change message.

DETAILED DESCRIPTION OF THE INVENTION

[0019] Figure 1 is a block diagram illustrating a typical system configuration for a storage area network. As illustrated, the overall system includes a first host device 108A and a second host device 108B. These host devices are typically computers or application servers of well known design. The hosts are coupled to a storage system 109, typically made up of a disk array, for example configured in accordance with a RAID protocol. The disk array 109 typically includes a large number of disk drives 106, of which one is illustrated. Disk 106 has been configured to have two volumes 105A and 105B.

[0020] The hosts 108 and storage system 109 are coupled together using a suitable means for exchanging data between them. A data link 110 is illustrated in Figure 1. The data link 110 can take any appropriate format, for example, a fibre channel network, a local area network, the internet, a private network, a network switch device, or any other interconnection means.

[0021] In a typical storage system 109, large numbers of storage volumes such as 105A and 105B will be provided. The storage volumes 105 themselves, or even portions thereof, are given logical unit numbers (LUNs) or other identification to identify them and/or provide addressing information for the addressing of information to be stored in those storage volumes, or retrieved from those storage volumes. The storage system 109 and the hosts 108 each include an interface 107 to provide an interface to the network. Interface 107 conventionally is provided by using a host bus adapter (HBA) or a gigabit Ethernet card, or other known interface. The selection of any particular interface will depend primarily upon the configuration of the data link 110 to which it is coupled.

[0022] Each of the hosts includes cluster management software 102 which operates on that host. The cluster management software in each host is coupled to similar software in other hosts. For example, as shown in Figure 1, the cluster management software 102A operating on host device 108A is coupled to communicate with the cluster management software 102B operating on host 108B. The cluster management software communicates with the corresponding software on other hosts to determine the ownership of storage volumes 105 for a particular host. Each storage volume 105 can be accessed by one or more computers that have the ownership for the volume. During initialization of the system, the hosts receive the LUNs, and, either independently, or with support from a system operator, the hosts decide which LUNs are to be associated with each host. After this has been

determined, the volume security control software 101 requests the storage device 109 to allow I/O accesses from that host to the designated volumes 105. Once the system has been appropriately initialized, then the security management program 103 in storage system 109 controls access to the storage volumes 105. The security management program stores the security configuration information 104 in a protected area for later use.

[0023] The system illustrated in Figure 1 is often termed an "in-band" control system. Figure 2 illustrates a different control system referred to as an "out-of-band" control system. The storage system 109 is the same in Figure 2 as that described in Figure 1. In contrast, however, the architecture of the system illustrated in Figure 2 places the responsibility for volume security control on a storage manager 112. Storage manager 112 is coupled to each of the hosts, and is responsible for the volume security control software 101. This software provides the same functionality as the volume security control software 101 provided in the system configuration shown in Figure 1. The volume security control software 101 running on the storage manager 112 coordinates management of the volume security. It also monitors the storage system and stores information about the configuration of the storage system and the like in local storage 111.

[0024] Figure 3 is a diagram illustrating the data structure for storage configuration information 111 (shown in Figure 2). This diagram illustrates how access is controlled to various portions of the storage. In Figure 3 to provide an example, volumes 2, 3 and 4 are shown in the column "Volume ID." Each volume is allocated to one or more ports to be activated. As shown by the diagram, volumes 2 and 3 are allocated to port 0, while volume 4 is allocated to port 1. The storage ID shown in Figure 3 is a storage identification parameter. This parameter identifies the storage asset, usually by serial number, node World Wide Name, or vendor-specific identification number. The node World Wide Name is a fixed, unique address that is assigned to disk array device 109. In Figure 3 the storage ID is shown as "Array #0" representing the first disk array associated with the storage product. The interface identification number (port 0 or port 1) represents the unique identification associated with a given network interface. Alternatively, the port WWN, or Port ID which are assigned to each port can be used to provide this information. The port WWN is a fixed, unique address that is assigned to each port. The port ID is an unique identifier that is assigned to each network interface hardware. Also, IP address or MAC address can be used if I/F 107 is a Ethernet card.

[0025] The host identification ("Host ID" in Figure 3) represents an identifier to designate a host that is permitted or restricted to access the particular designated volume (in that row of the table). The LUN security function uses the node WWN of the host, or the port WWN of the HBA, or the MAC address of the network card installed on the host to provide this identification parameter. To illustrate its operation, Figure 3 shows the hypothetical example that volume 3 on port 0 of array 0 is allowed to be accessed by hosts 8 and 9, but access is not permitted for host 12.

[0026] Figure 4 is a diagram illustrating an example of the data structure for the security configuration information 104 found in disk array unit 109. As shown by Figure 4, this data structure is synchronized with, and corresponds to, the storage configuration information 111 from the management unit 112. As also illustrated, the storage ID is not required since this information is maintained in the storage unit itself (as identified by Figure 3).

[0027] Figure 5 is a flowchart illustrating a preferred embodiment of the method of operation of this invention. The diagram shown in Figure 5 is divided into two parts -- operations that occur within the host (on the left of the diagram), and operations that occur within the disk array (skewed to the right of the diagram). The process begins with step 501 in the host in which cluster management software negotiates to determine the ownership or control of the disk resources. With reference to Figure 1, this step is carried out by the software 102A in host 108A negotiating with the software 102B in host 108B. Such a negotiation typically uses the known SCSI-3 persistent reserve algorithm, or the SCSI-2 challenge/defense protocol. At the conclusion of the process, the host computers will have rights to access particular LUNs (or volumes) in the disk array.

[0028] Step 502 in Figure 5 illustrates that the negotiation concludes with the cluster management software notifying the volume security control software of the results of the negotiation. The volume security control software 101, as described in conjunction with Figures 1 and 2, will reside either in each of the hosts (Figure 1) or in a manager (Figure 2).

[0029] Figure 6 is an illustration of an ownership change message sent by the cluster management software 102 to the volume security control software 101. As shown in Figure 6, the hypothetical message illustrates that volume number 2 has been acquired by host number 8, and that volume number 3 has been lost to host number 8.

[0030] Returning to the process illustrated in Figure 5, after sending the message, at step 503 the volume security control software 101 will request the disk array 109 to change the configuration for volume security. This is carried out by the volume security control software 503 sending a request message to the security management program 103 found in the disk array 109. This message requests changes in the LUN (or volume) security configuration.

[0031] Figure 7 illustrates a typical message sent by the volume security control software 101 to the security management program 103. As shown in Figure 7, an access control status change message is being transferred to illustrate that host number 8 is now permitted to access volume number 2 and is no longer permitted to access volume number 3.

[0032] Again returning to Figure 5, once the message of Figure 7 is received by the security management program 103, it carries out step 504 in Figure 5. At this time the security management program 103 reconfigures the LU security settings and updates the security configuration information 104. This operation will result in new entries in the security configuration information table (shown in Figure 4) by which the status for volume number 2 and host number 8 is changed from "deny" to "permit." Similarly, the status for volume number 3 and host number 8 is switched from "permit" to "deny."

[0033] Following the disk array device operation of step 504, the host device carries out step 505. In this step the cluster management software maintains control of the disk resources to verify that no inconsistent status has occurred. This is typically carried out using a "heartbeat" communication protocol. As shown by step 506 in Figure 5, if the heartbeat is lost, or if an inconsistent status is detected, then the cluster management software 102 will reset the system and restart the negotiation process to allocate disk resources to hosts.

[0034] The foregoing has been a description of preferred embodiments of this invention. It should be appreciated that departures from the specific embodiment illustrated may be made while remaining within the scope of this invention. For example security for the storage devices may be implemented on the basis of other than LUs or volumes, instead using addresses or other designations. The scope of the invention is defined by the appended claims.